Moderate Deviations for Queues in Critical Loading

Anatolii A. Puhalskii
Institute for Problems
in Information Transmission
19 Bolshoi Karetnii
101447 Moscow Russia

July 29, 1997

Abstract

We establish logarithmic asymptotics of moderate deviations for the processes of queue length and waiting times in single server queues and open queueing networks in critical loading. Our results complement earlier heavy-traffic approximation results.

Keywords and phrases: large deviation principle, queues, heavy traffic

AMS subject classification:

primary: 60F10, 60K25

1. Introduction

This paper complements the classical results on heavy-traffic approximation for queues in Kingman [7], Prohorov [12], Iglehart and Whitt [5], Borovkov [1], and Reiman [19] by studying some related large-deviation asymptotics. In a standard set-up, one considers a sequence of $GI/GI/1$ queues indexed by n with associated loads $\rho_n \to 1$ as $n \to \infty$ so that

$$\sqrt{n}(1-\rho_n) \to c \text{ where } |c| < \infty, \tag{1.1}$$

and establishes convergence in distribution of suitably time-scaled and normalized queue-related processes to processes of diffusion type; say, the processes $(n^{-1/2}W_n(nt), t \geq 0)$, where $W_n(t)$ denotes the unfinished work (or virtual waiting time) in the n th system at time t, converge in distribution in the Skorohod J_1 topology to a reflected Brownian motion with drift [5]. The limits when $|\sqrt{n}(1-\rho_n)| \to \infty$ are different: in the context of the unfinished work again, if $\sqrt{n}(1-\rho_n) \to \infty$, then the $n^{-1/2}W_n(nt)$ converge to 0 in probability, and if $\sqrt{n}(1-\rho_n) \to -\infty$, a proper limit have the processes $\left(n^{-1/2}(W_n(nt) - (\rho_n - 1)nt), t \geq 0\right)$. Our focus here is on large deviation asymptotics for the latter case: $|\sqrt{n}(1-\rho_n)| \to \infty$. More specifically, we assume that, for some $b_n \to \infty$ with $b_n = o(\sqrt{n})$, we have that

$$\frac{1}{b_n}\sqrt{n}(1-\rho_n) \to c \text{ where } |c| < \infty, \tag{1.2}$$

and study the logarithmic asymptotics of the large deviations of processes like $(b_n^{-1}n^{-1/2}W_n(nt), t \geq 0)$ as $n \to \infty$. In Wentzell's classification of large deviations, Wentzell [23], this is the case of "moderate deviations" since the choice $b_n = 1$ specifies "normal deviations", and $b_n = \sqrt{n}$, "very large deviations". ¿From an application viewpoint, we are concerned with the queue behaviour at times much greater than $(1-\rho)^{-2}$ for ρ close to 1 while the standard heavy-traffic results refer to time intervals of order $(1-\rho)^{-2}$. Accordingly, to distinguish from standard heavy traffic, we refer to the regime specified by condition (1.2) as *near-heavy traffic*.

We now give an outline of the paper and a summary of the results. Section 2 contains technical preliminaries. In Section 3 we consider FIFO single server queues in near-heavy traffic. Section 4 extends the results to the case of FIFO open queueing networks with homogeneous customer population. The results mostly have the form of large deviation principles (LDPs) in the spaces of right-continuous functions with left limits equipped with one of Skorohod's topologies, Skorohod [21], for such processes as the processes of queue length, unfinished work, completed work,

waiting times, the number of departures, and departure times. Occasionally, we give LDPs for one-dimensional projections. The rate functions that we obtain are quadratic in form and reminiscent of the distributions of the diffusion processes arising in the corresponding heavy-traffic limit theorems. Moreover, the ideas of the proofs are either borrowed from the proofs of the corresponding weak convergence results or could be used to give them alternative proofs. So, in a sense, the paper is another evidence of the analogy between large deviation theory and weak convergence theory, Puhalskii [13]–[16].

2. Technical Preliminaries

We shall work in the function space $D(R^d) \equiv D([0,\infty), R^d)$ of right-continuous R^d-valued functions on $[0,\infty)$ with left limits, endowed with the Skorohod [21] J_1 or M_1 topologies, or a modification of the M_1 topology denoted by M_1', we refer to [8], [11], [24], [9], [17] for details. These spaces are metrizable as separable metric spaces and have Borel σ-fields coinciding with the σ-field generated by the coordinate projections. For $x = (x(t), t \geq 0) \in D(R^d)$, we denote by $x(t-)$ the left limit at t and by $\Delta x(t)$, the jump at t: $\Delta x(t) \equiv x(t) - x(t-), t > 0$; $\Delta x(0) \equiv x(0)$.

As in Varadhan [22], we say that a function $I(x)$ defined on a metric space S and taking values in $[0,\infty]$ is *a rate function* if the sets $\{x \in S : I(x) \leq a\}$ are compact for all $a \geq 0$, and a sequence $\{P_n, n \geq 1\}$ of probability measures on the Borel σ-field of S (or a sequence of random elements $\{X_n, n \geq 1\}$ with values in S and distributions P_n) obeys a large deviation principle (LDP) for a normalizing sequence $a_n \to \infty$ with the rate function I if

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log P_n(F) \leq - \inf_{x\in F} I(x) \tag{2.1}$$

for all closed $F \subset S$, and

$$\underline{\lim_{n\to\infty}} \frac{1}{a_n} \log P_n(G) \geq - \inf_{x\in G} I(x) \tag{2.2}$$

for all open $G \subset S$.

The standard choice of the normalizing sequence is $a_n = n$. We mention that an LDP can always be reduced to this standard form by reparametrizing the family $\{P_n, n \geq 1\}$, Varadhan [22], however allowing a general normalizing sequence seems more convenient in applications. We refer to [20], [22] and [17] for additional background.

We say that a sequence $\{X_n, n \geq 1\}$ of random elements of a metric space (S, ρ) converges *super-exponentially in probability at rate* a_n to an element $x_0 \in S$ if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P^{1/a_n}(\rho(X_n, x_0) > \epsilon) = 0 \tag{2.3}$$

and we write $X_n \overset{P^{1/a_n}}{\to} x_0$. This mode of convergence plays a role in large deviations similar to the role convergence in probability plays in weak convergence. Properties of super-exponential convergence in probability which we invoke below can be found in Puhalskii and Whitt [17]. Here we just mention that if (2.3) holds for $S = D(R^d)$ with one of the topologies J_1, M_1 or M_1' and x_0 is a continuous function starting at 0, then it also holds for the locally uniform metric on $D(R^d)$, [17, Lemma 4.2(a)].

The following easy consequence of the contraction principle [22] comes in handy below.

Lemma 2.1. *Let X_n, Y_n and Z_n be random variables with values in metric spaces S_X, S_Y and S_Z, respectively, and let $S \equiv S_X \times S_Y \times S_Z$ be endowed with product topology. Assume that the sequence $\{(X_n, Y_n, Z_n), n \geq 1\}$ obeys an LDP in S for a normalizing sequence a_n with rate function $I_{X,Y,Z}(x,y,z)$ which is finite only if $y = f(x)$, where $f : X \to Y$ is a bijection. Then the sequences $\{(X_n, Z_n), n \geq 1\}$ and $\{(Y_n, Z_n), n \geq 1\}$ obey LDPs with the respective rate functions $I_{X,Z}$ and $I_{Y,Z}$ given by the equalities*

$$I_{X,Z}(x,z) = I_{Y,Z}(f(x),z) = I_{X,Y,Z}(x,f(x),z),$$

$$I_{Y,Z}(y,z) = I_{X,Z}(f^{-1}(y),z) = I_{X,Y,Z}(f^{-1}(y),y,z).$$

When dealing with an LDP for stationary waiting times, we will use the following version of Lemma 4.1 in Puhalskii [15].

Lemma 2.2. *Let $\{P_n, n \geq 1\}$ be a sequence of probability measures on R and I be a rate function on R. If the bounds (2.1) and (2.2) hold for the sets $F = [a,b], [a,\infty), (-\infty, b]$ and $G = (a,b)$, where $a, b \in R$, then $\{P_n, n \geq 1\}$ obeys an LDP with the rate function I.*

The proof follows by Lemma 4.1 in [15].

The next easy lemma is a consequence of the extended contraction principle [17] and continuity of the supremum map, Whitt [24, Theorems 6.2 and 6.3], the latter theorems carrying over to the M_1' topology (see the argument of the proof of Theorem 5.1 in [17] for more detail).

Lemma 2.3. *Let $X_n = (X_n(t), t \geq 0)$ be random processes with paths from $D(R^d)$. If the sequence $\{X_n, n \geq 1\}$ obeys an LDP in $D(R^d)$ for one of the topologies J_1, M_1 or M_1' and normalizing sequence a_n with rate function $I_X(x)$ which equals infinity at elements of $D(R^d)$ that are either discontinuous or not equal to 0 at 0, then, for every $\varepsilon > 0$ and $t > 0$,*

$$\lim_{n\to\infty} P^{1/a_n}(\sup_{0\leq s\leq t} |\Delta X_n(s)| > \varepsilon) = 0.$$

Rate functions in the limit theorems below are generally defined in terms of solutions to Skorohod problems with skew reflection [4, 19, 3]. We now recall the relevant definitions. Let $P = (p_{kl})$

be a $K \times K$ matrix with nonnegative entries and spectral radius less than unity. Denote by P^T the transpose of P and let $\mathcal{R}_P$ denote the map from $D(R^K)$ into $D(R^K)$ associating to each $x = (x(t), t \geq 0) \in D(R^K)$ with $x_k(0) \geq 0, 1 \leq k \leq K$, the function $z = (z(t), t \geq 0) \in D(R^K)$ such that

1. $z = x + (I - P^T)y$,

2. y is componentwise nondecreasing with $y_k(0) = 0, 1 \leq k \leq K$,

3. $z_k(t) \geq 0$ and $\int_0^\infty z_k(t)\, dy_k(t) = 0$, $1 \leq k \leq K$.

The map $\mathcal{R}_P$ is well defined and Lipshitz continuous for the locally uniform metric on $D(R^K)$, Harrison and Reiman [4], Reiman [19], Mandelbaum [10], Chen and Mandelbaum [2]. In the one-dimensional case $K = 1$ and $P = (0)$ the reflection map which we denote $\mathcal{R}$ has the explicit form

$$\mathcal{R}(x)(t) = x(t) - \inf\{x(s) : 0 \leq s \leq t\} \wedge 0, \ t \geq 0 \ . \tag{2.4}$$

The following characterization of skew reflection is in the spirit of Lemma 3.1 in [18] and Lemma 4.6 in [15], and proved by the same argument.

Lemma 2.4. *Let $z \in D(R^K)$ be componentwise nonnegative and $x \in D(R^K)$ be componentwise absolutely continuous. Then $z = \mathcal{R}_P(x)$ if and only if z is absolutely continuous and there exists an absolutely continuous function $y \in D(R^K)$ with the properties*

$$\dot{z}(t) = \dot{x}(t) + (I - P^T)\dot{y}(t) \ a.e.$$

and

$$y_k(0) = 0, \ \dot{y}_k(t) \geq 0 \ a.e., \ z_k(t)\dot{y}_k(t) = 0 \ a.e., \ 1 \leq k \leq K.$$

Thus $\dot{z}(t)$ a.e. solves a linear complementarity problem [10, 2].

3. Moderate Deviations for Single Server Queues in Near-Heavy Traffic

We consider a sequence of FIFO single server queues indexed by n. We assume that the queues are initially empty. Let $A_n(t)$ denote the number of arrivals by t, $S_n(t)$, the number of customers served for the first t units of the server's busy time, $D_n(t)$, the number of departures by t, $Q_n(t)$, the queue length at t, $W_n(t)$, the unfinished work at t, $C_n(t)$, the completed work at t, $H_n(k)$, the waiting time of the k th customer, and $L_n(k)$, the departure time of the k th customer.

Let also

$$V_n(k) \equiv \min\{t : S_n(t) \geq k\},\ V_n(0) = 0, \tag{3.1}$$

be the cumulative service time of the first k customers.

Denoting by $\circ$ the composition map, we have the following obvious equalities

$$W_n(t) = V_n \circ A_n(t) - C_n(t), \tag{3.2}$$

$$C_n(t) = \int_0^t 1(W_n(s) > 0)\, ds = \int_0^t 1(Q_n(s) > 0)\, ds, \tag{3.3}$$

$$Q_n(t) = A_n(t) - D_n(t), \tag{3.4}$$

$$D_n(t) = S_n \circ C_n(t), \tag{3.5}$$

Let $b_n \to \infty$ and $b_n/\sqrt{n} \to 0$ as $n \to \infty$, and λ_n and μ_n be positive numbers. We define the associated normalized and time-scaled processes by

$$\overline{A}_n = (\overline{A}_n(t), t \geq 0),\ \overline{A}_n(t) = \frac{1}{b_n\sqrt{n}}(A_n(nt) - \lambda_n nt), \tag{3.6}$$

$$\overline{S}_n = (\overline{S}_n(t), t \geq 0),\ \overline{S}_n(t) = \frac{1}{b_n\sqrt{n}}(S_n(nt) - \mu_n nt), \tag{3.7}$$

$$\overline{V}_n = (\overline{V}_n(t), t \geq 0),\ \overline{V}_n(t) = \frac{1}{b_n\sqrt{n}}(V_n(\lfloor nt \rfloor) - \mu_n^{-1} nt), \tag{3.8}$$

$$\overline{D}_n = (\overline{D}_n(t), t \geq 0),\ \overline{D}_n(t) = \frac{1}{b_n\sqrt{n}}(D_n(nt) - \mu_n nt), \tag{3.9}$$

$$\overline{W}_n = (\overline{W}_n(t), t \geq 0),\ \overline{W}_n(t) = \frac{1}{b_n\sqrt{n}} W_n(nt), \tag{3.10}$$

$$\overline{Q}_n = (\overline{Q}_n(t), t \geq 0),\ \overline{Q}_n(t) = \frac{1}{b_n\sqrt{n}} Q_n(nt), \tag{3.11}$$

$$\overline{C}_n = (\overline{C}_n(t), t \geq 0),\ \overline{C}_n(t) = \frac{1}{b_n\sqrt{n}} (C_n(nt) - nt), \tag{3.12}$$

$$\overline{H}_n = (\overline{H}_n(t), t \geq 0),\ \overline{H}_n(t) = \frac{1}{b_n\sqrt{n}} H_n(\lfloor nt \rfloor + 1), \tag{3.13}$$

$$\overline{L}_n = (\overline{L}_n(t), t \geq 0),\ \overline{L}_n(t) = \frac{1}{b_n\sqrt{n}}(L_n(\lfloor nt \rfloor + 1) - \mu_n^{-1} nt). \tag{3.14}$$

We assume that $\lambda_n \to \lambda > 0$ and $\mu_n \to \mu > 0$ as $n \to \infty$, and the near-heavy traffic condition holds:

$$\frac{1}{b_n}\sqrt{n}(\lambda_n - \mu_n) \to r,\ -\infty < r < \infty. \tag{3.15}$$

Note that (3.15) implies that $\lambda = \mu$.

The next theorem parallels the results of Iglehart and Whitt [5], on the one hand, and Theorems 3.1 and 4.1 in Puhalskii and Whitt [18], on the other hand. Let us denote $e = (t, t \geq 0)$.

Theorem 3.1.

(a) *Assume that* $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for one of the topologies* J_1, M_1 *or* M_1' *and normalizing sequence* b_n^2 *with rate function* $I_{A,S}(a,s)$. *Then* $\{(\overline{Q}_n, \overline{D}_n, \overline{C}_n), n \geq 1\}$ *obeys an LDP in* $D(R^3)$ *for the same topology and normalizing sequence* b_n^2 *with rate function*

$$I_{Q,D,C}(q,d,c) = \inf_{\substack{a,s \in D(R^2): \\ q=\mathcal{R}(a-s+re),\, d=a-q+re, \\ c=\mu^{-1}(a-s-q+re)}} I_{A,S}(a,s) \ .$$

(b) *Assume that* $\{(\overline{A}_n, \overline{V}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for one of the topologies* J_1, M_1 *or* M_1' *and normalizing sequence* b_n^2 *with rate function* $I_{A,V}(a,v)$. *Then* $\{(\overline{W}_n, \overline{C}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for the same topology and normalizing sequence* b_n^2 *with rate function*

$$I_{W,C}(w,c) = \inf_{\substack{a,v \in D(R^2): \\ w=\mathcal{R}(v \circ (\mu e)+\mu^{-1}a+\mu^{-1}re), \\ c=v \circ (\mu e)+\mu^{-1}a+\mu^{-1}re-w}} I_{A,V}(a,v) \ .$$

(c) *Assume that* $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for one of the topologies* J_1 *or* M_1' *and normalizing sequence* b_n^2 *with rate function* $I_{A,S}(a,s)$, *which, in the case of the* J_1 *topology, is infinite when* s *is either discontinuous or not equal to* 0 *at* 0. *Then* $\{(\overline{Q}_n, \overline{D}_n, \overline{W}_n, \overline{C}_n), n \geq 1\}$ *obeys an LDP in* $D(R^4)$ *for the same topology and normalizing sequence* b_n^2 *with rate function*

$$I_{Q,D,W,C}(q,d,w,c) = \inf_{\substack{a,v \in D(R^2): \\ q=\mathcal{R}(a-s+re),\, d=a-q+re, \\ w=\mu^{-1}\mathcal{R}(a-s+re), \\ c=\mu^{-1}(a-s+re)-w}} I_{A,S}(a,s) \ .$$

(d) *Assume that* $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for one of the topologies* J_1, M_1 *or* M_1' *and normalizing sequence* b_n^2 *with rate function* $I_{A,S}(a,s)$, *which is infinite when either* a *or* s *is either discontinuous or not equal to* 0 *at* 0. *Then the sequence* $\{(\overline{Q}_n, \overline{D}_n, \overline{W}_n, \overline{C}_n, \overline{H}_n, \overline{L}_n), n \geq 1\}$ *obeys an LDP in* $D(R^6)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function* $I_{Q,D,W,C,H,L}(q,d,w,c,h,l) = I_{Q,D,C}(q,d,c)$ *when* $q = \mu w$, $h = w \circ (\mu^{-1}e)$ *and* $d = -\mu l \circ (\mu e)$, *and* $I_{Q,D,W,C,H,L}(q,d,w,c,h,l) = \infty$ *otherwise.*

Proof. We begin with a proof of (a). By (3.4), (3.5), (3.3), (3.11), (3.6) and (3.7),

$$\overline{Q}_n(t) = \overline{A}_n(t) - \overline{S}_n \circ \overline{C}_n'(t) + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)t + \frac{\sqrt{n}}{b_n}\mu_n \int_0^t 1(\overline{Q}_n(s) = 0)\, ds, \tag{3.16}$$

$$\overline{D}_n(t) = \overline{S}_n \circ \overline{C}_n'(t) + \mu_n \overline{C}_n(t), \tag{3.17}$$

$$\overline{C}_n(t) = -\frac{\sqrt{n}}{b_n} \int_0^t 1(\overline{Q}_n(s) = 0)\, ds, \tag{3.18}$$

where

$$\overline{C}'_n(t) = \frac{1}{n} C_n(nt) = \int_0^t 1(\overline{Q}_n(s) > 0)\, ds. \tag{3.19}$$

Since $\overline{Q}_n(t)$ is nonnegative and $\int_0^t 1(\overline{Q}_n(s) = 0)\, ds$ increases only when $\overline{Q}_n(t) = 0$, we conclude from (3.16) that the process $(\overline{Q}_n(t), t \geq 0)$ is the Skorohod reflection of the process $(\overline{A}_n(t) - \overline{S}_n \circ \overline{C}'_n(t) + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)t, t \geq 0)$:

$$\overline{Q}_n = \mathcal{R}\left(\overline{A}_n - \overline{S}_n \circ \overline{C}'_n + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)e\right), \tag{3.20}$$

and, by (3.18),

$$\mu_n \overline{C}_n = \overline{A}_n - \overline{S}_n \circ \overline{C}'_n + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)e - \overline{Q}_n. \tag{3.21}$$

By the Lipshitz property of the reflection in the locally uniform metric, we have, for some $K(t) > 0$,

$$\mu_n |\overline{C}_n(t)| \leq K(t) \sup_{s \leq t} \left| \overline{A}_n(s) - \overline{S}_n \circ \overline{C}'_n(s) + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)s \right|, \ t \geq 0. \tag{3.22}$$

The LDP for $(\overline{A}_n, \overline{S}_n)$, the inequality $\overline{C}'_n(t) \leq t$ and (3.15) imply that

$$\lim_{\alpha \to \infty} \overline{\lim_{n \to \infty}} P^{1/a_n}\left(\sup_{s \leq t} \left| \overline{A}_n(s) - \overline{S}_n \circ \overline{C}'_n(s) + \frac{\sqrt{n}}{b_n}(\lambda_n - \mu_n)s \right| > \alpha \right) = 0$$

with $a_n = b_n^2$. Hence, by (3.22) and (3.18), since $\sqrt{n}/b_n \to \infty$ and $\mu_n \to \mu > 0$,

$$\int_0^t 1(\overline{Q}_n(s) = 0)\, ds \overset{P^{1/a_n}}{\to} 0 \text{ as } n \to \infty, \ t > 0, \tag{3.23}$$

so by (3.19) and Lemma 3.1 in [14], (for the locally uniform metric on $D(R)$) $\overline{C}'_n \overset{P^{1/a_n}}{\to} e$, and an obvious extension of Lemma 4.3 in [17] implies by the LDP for $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ with $I_{A,S}$ that the sequence $\{(\overline{A}_n, \overline{S}_n \circ \overline{C}'_n), n \geq 1\}$ obeys an LDP with $I_{A,S}$. The required now follows by (3.20), (3.17), (3.21), continuity of the reflection, the near-heavy traffic condition (3.15), and the contraction principle.

The argument for parts (b) and (c) is similar. For (b), write by (3.2), (3.3), (3.6), (3.8), (3.10), and (3.12),

$$\overline{W}_n(t) = \overline{V}_n \circ \overline{A}'_n(t) + \mu_n^{-1}\overline{A}_n(t) + \frac{\sqrt{n}}{b_n}(\rho_n - 1)t + \frac{\sqrt{n}}{b_n}\int_0^t 1(\overline{W}_n(s) = 0)\, ds,$$

$$\overline{C}_n(t) = -\frac{\sqrt{n}}{b_n}\int_0^t 1(\overline{W}_n(s) = 0)\, ds,$$

where $\overline{A}'_n(t) = n^{-1}A_n(nt)$, and note that the LDP for $\{\overline{A}_n, n \geq 1\}$ implies by [17, Lemma 4.2(b)] that

$$\overline{A}'_n \overset{P^{1/a_n}}{\to} \mu e. \tag{3.24}$$

Part (c) follows by combining the preceding arguments if one notes that by (3.1) and Theorem 5.4 in [17] the assumptions imply that the sequence $\{(\overline{A}_n, \overline{S}_n, \overline{V}_n), n \geq 1\}$ obeys an LDP in $D(R^3)$ for one of the topologies J_1 or M_1' and normalizing sequence b_n^2 with rate function $I_{A,S,V}(a,s,v) = I_{A,S}(a,s)$, when $s = -\mu v \circ (\mu e)$, and $I_{A,S,V}(a,s,v) = \infty$ otherwise.

We now prove (d). Since the rate function $I_{A,S}(a,s)$ equals infinity at elements of $D(R^2)$ that are either discontinuous or not equal to 0 at 0, the extended contraction principle [17] implies that under the assumptions $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ obeys an LDP in $D(R^2)$ for the J_1 topology with the rate function $I_{A,S}(a,s)$.

Let

$$U_n(k) = \inf\{t \geq 0 : A_n(t) \geq k\} \tag{3.25}$$

and $\overline{U}_n'(t) = U_n(\lfloor nt \rfloor + 1)/n$. By (3.24) and Lemma 4.2(c) in [17],

$$\overline{U}_n' \overset{P^{1/a_n}}{\to} \mu^{-1} e. \tag{3.26}$$

Noting also that $L_n(k) = \inf\{t \geq 0 : D_n(t) \geq k\}$, we conclude by Lemma 4.3 and Theorem 5.4 in [17], and part (c) of the theorem we are proving that $\{(\overline{Q}_n, \overline{D}_n, \overline{W}_n, \overline{C}_n, \overline{W}_n \circ \overline{U}_n', \overline{L}_n), n \geq 1\}$ obeys an LDP in $D(R^6)$ for the J_1 topology with rate function $I'(q,d,w,c,h,l) = I_{Q,D,C}(q,d,c)$ when $q = \mu w$, $h = w \circ (\mu^{-1} e)$ and $d = -\mu l \circ (\mu e)$, and $I'(q,d,w,c,h,l) = \infty$ otherwise.

We now prove that

$$\overline{H}_n - \overline{W}_n \circ \overline{U}_n' \overset{P^{1/a_n}}{\to} 0 \tag{3.27}$$

which will conclude the proof by Lemma 4.1(c) in [17]. Since $\overline{W}_n(\overline{U}_n'(t)-) \leq \overline{H}_n(t) \leq \overline{W}_n(\overline{U}_n'(t))$, we have that

$$\sup_{s \leq t} |\overline{H}_n(s) - \overline{W}_n \circ \overline{U}_n'(s)| \leq \sup_{0 \leq s \leq \overline{U}_n'(t)} |\Delta \overline{W}_n(s)|. \tag{3.28}$$

Since $I_{A,S}(a,s)$ equals infinity when either one of the arguments is either a discontinuous function or not equal to 0 at 0, part (c) of the theorem implies that $\{\overline{W}_n, n \geq 1\}$ obeys an LDP in $D(R)$ for the J_1 topology with rate function which equals infinity both at discontinuous functions from $D(R)$ and functions not equal to 0 at 0 so that by Lemma 2.3, for $t > 0$,

$$\sup_{0 \leq s \leq t} |\Delta \overline{W}_n(s)| \overset{P^{1/a_n}}{\to} 0 \tag{3.29}$$

as $n \to \infty$. Putting together (3.26), (3.28) and (3.29) proves (3.27). The theorem is proved.

Remark 3.1. Let $I_n(t)$ denote the cumulative server's idle time at t, i.e., $I_n(t) = \int_0^t 1(Q_n(s) = 0)ds$, and $\overline{I}_n(t) = I_n(nt)/(b_n\sqrt{n})$. Since obviously $\overline{I}_n(t) = -\overline{C}_n(t)$ (see (3.18)), the theorem provides LDPs for $\{\overline{I}_n, n \geq 1\}$ as well.

Remark 3.2. Parts (c) and (d) show that under the hypotheses "Little's law" holds: if the rate function is finite, then $\mu w = q$. So, (Q, D, C) is "a sufficient statistic" in the sense of Lemma 2.1.

We now consider the case of quadratic rate functions typical of the LDP for partial sums of triangular arrays of i.i.d. sequences (see Lemma 6.1 of [17] or [14, Example 7.2]) or partial sums of interarrival times in superpositions of renewal processes (see Theorem 7.2 of [17]). We adopt the convention $0/0 = 0$ so that, e.g., the rate function $I_A(a)$ below, in the case if $\sigma_A = 0$, equals 0 when $a(t) = 0$ for all $t \geq 0$ and equals ∞ otherwise.

Theorem 3.2. *Let condition (3.15) hold. Assume that $\{(\overline{A}_n, \overline{S}_n), n \geq 1\}$ obeys an LDP in $D(R^2)$ for the J_1 topology and normalizing sequence b_n^2 with rate function*

$$I_{A,S}(a, s) = I_A(a) + I_S(s),$$

where

$$I_A(a) = \frac{1}{2\sigma_A^2} \int_0^\infty \dot{a}(t)^2 dt \tag{3.30}$$

for a absolutely continuous with $a(0) = 0$ and $I_A(a) = \infty$ otherwise, and

$$I_S(s) = \frac{1}{2\sigma_S^2} \int_0^\infty \dot{s}(t)^2 dt \tag{3.31}$$

for s absolutely continuous with $s(0) = 0$ and $I_S(s) = \infty$ otherwise. Then the following holds.

(a) *The sequence $\{(\overline{Q}_n, \overline{D}_n, \overline{C}_n), n \geq 1\}$ obeys an LDP in $D(R^3)$ for the J_1 topology and normalizing sequence b_n^2 with rate function*

$$I_{Q,D,C}(q, d, c) = \int_0^\infty 1(q(t) > 0) \left[\frac{1}{2\sigma_A^2} (\dot{q}(t) + \dot{d}(t) - r)^2 + \frac{1}{2\sigma_S^2} \dot{d}(t)^2 \right] dt$$
$$+ \int_0^\infty 1(q(t) = 0) \left[\frac{1}{2\sigma_A^2} (\dot{d}(t) - r)^2 + \frac{1}{2\sigma_S^2} (\dot{d}(t) - \mu \dot{c}(t))^2 \right] dt,$$

when q, d and c are absolutely continuous with $q(0) = d(0) = c(0) = 0$, q is nonnegative, c is nonpositive and nonincreasing, $\dot{c}(t) = 0$ a.e. on the set $q(t) > 0$, and $I_{Q,D,C}(q, d, c) = \infty$ otherwise.

(b) *The sequence $\{(\overline{Q}_n, \overline{D}_n), n \geq 1\}$ obeys an LDP in $D(R_\uparrow^2)$ for the J_1 topology and normalizing sequence b_n^2 with rate function*

$$I_{Q,D}(q, d) = \int_0^\infty 1(q(t) > 0) \left[\frac{1}{2\sigma_A^2} (\dot{q}(t) + \dot{d}(t) - r)^2 + \frac{1}{2\sigma_S^2} \dot{d}(t)^2 \right] dt$$
$$+ \int_0^\infty 1(q(t) = 0) \left[\frac{1}{2\sigma_A^2} (\dot{d}(t) - r)^2 + \frac{1}{2\sigma_S^2} 1(\dot{d}(t) > 0) \dot{d}(t)^2 \right] dt,$$

when q and d are absolutely continuous with $q(0) = d(0) = 0$, q is nonnegative, and $I_{Q,D}(q,d) = \infty$ otherwise.

(c) *The sequence $\{(\overline{Q}_n, \overline{C}_n), n \geq 1\}$ obeys an LDP in $D(R^2)$ for the J_1 topology and normalizing sequence b_n^2 with rate function*

$$I_{Q,C}(q,c) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(q(t) > 0)(\dot{q}(t) - r)^2 \, dt + \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(q(t) = 0)(\mu \dot{c}(t) - r)^2 \, dt,$$

when q and c are absolutely continuous with $q(0) = c(0) = 0$, q is nonnegative, c is nonpositive and nonincreasing, $\dot{c}(t) = 0$ a.e. on the set $q(t) > 0$, and $I_{Q,C}(q,c) = \infty$ otherwise.

(d) *The sequence $\{\overline{Q}_n, n \geq 1\}$ obeys an LDP in $D(R)$ for the J_1 topology and normalizing sequence b_n^2 with rate function*

$$I_Q(q) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(q(t) > 0)(\dot{q}(t) - r)^2 \, dt + \frac{1(r > 0) r^2}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(q(t) = 0) \, dt,$$

when q is nonnegative and absolutely continuous with $q(0) = 0$, and $I_Q(q) = \infty$ otherwise.

(e) *The sequence $\{\overline{C}_n, n \geq 1\}$ obeys an LDP in $D(R)$ for the J_1 topology and normalizing sequence b_n^2 with rate function I_C which is as follows. Let $k(c) = \text{ess sup}\, \{t > 0 : \dot{c}(t) < 0\}$.*

If $r < 0$, then

$$I_C(c) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty (\mu \dot{c}(t) - r)^2 \, dt,$$

when c is absolutely continuous, $c(0) = 0$, $\dot{c}(t) \leq 0$ a.e. and $k(c) = \infty$, and $I_C(c) = \infty$ otherwise.

If $r \geq 0$, then

$$I_C(c) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^{k(c)} (\mu \dot{c}(t) - r)^2 \, dt,$$

when c is absolutely continuous, $c(0) = 0$, $\dot{c}(t) \leq 0$ a.e., and $I_C(c) = \infty$ otherwise.

Proof. An application of Theorem 3.1(a), Lemma 2.4 (or Lemma 3.1 in [18]) and Lemma 3.3 in [15] yields the rate function of part (a). The rate functions in (b)–(e) follow by the contraction principle. In particular, in part (e) it can be proved in analogy with the proof of Theorem 5.1(b) in [18] that $\inf_q I_{Q,C}(q,c)$ is attained at $q(t) = 0$ for $t < k(c)$.

Remark 3.3. Let

$$\overline{U}_n(t) = \frac{U_n(\lfloor nt \rfloor) - \mu_n^{-1} nt}{b_n \sqrt{n}}$$

where $U_n(k)$ is defined by (3.25) and $U_n(0) = 0$. By (3.1) and an easy extension of Theorem 5.4 in [17] to the multidimensional case, the assumed LDP for $(\overline{A}_n, \overline{S}_n)$ holds if and only if the sequence $\{(\overline{U}_n, \overline{V}_n), n \geq 1\}$ obeys an LDP in $D(R^2)$ for the J_1 topology and normalizing sequence b_n^2 with rate function $I_{U,V} = I_U + I_V$, where

$$I_U(u) = \frac{1}{2\sigma_U^2} \int_0^\infty \dot{u}(t)^2 dt$$

for u absolutely continuous with $u(0) = 0$ and $I_U(u) = \infty$ otherwise, and

$$I_V(v) = \frac{1}{2\sigma_V^2} \int_0^\infty \dot{v}(t)^2 dt$$

for v absolutely continuous with $v(0) = 0$ and $I_V(v) = \infty$ otherwise, and $\sigma_U^2 = \sigma_A^2/\lambda^3$ and $\sigma_V^2 = \sigma_S^2/\mu^3$.

More specifically, for a $GI/GI/1$ queue, i.e., when A_n and S_n are renewal processes, let us denote by u_n the generic interarrival time and by v_n the generic service time. Then the LDP for $(\overline{U}_n, \overline{V}_n)$ holds if

$$\lambda_n^{-1} = Eu_n,\ \mu_n^{-1} = Ev_n,$$

$$\text{Var } u_n \to \sigma_U^2,\ \text{Var } v_n \to \sigma_V^2,$$

and either one of the following conditions is met:

(i) $\sup_n E(u_n)^{2+\epsilon} < \infty$, $\sup_n E(v_n)^{2+\epsilon} < \infty$ for some $\epsilon > 0$ and $\sqrt{\log n}/b_n \to \infty$;

(ii) $\sup_n E\exp(\alpha u_n^\beta) < \infty$, $\sup_n E\exp(\alpha v_n^\beta) < \infty$ for some $\alpha > 0, 0 < \beta \leq 1$ and $n^{\beta/2}/b_n^{2-\beta} \to \infty$.

This follows by Lemma 6.1 and Theorem 5.4 in [17].

Remark 3.4. It is interesting to compare I_C with the rate function for the arrived work. Since under the conditions of the theorem $\{(A_n, V_n), n \geq 1\}$ obeys an LDP with rate function $I_{A,V} = I_A + I_V$, it easily follows that the processes $((V_n \circ A_n(nt) - nt)/(b_n\sqrt{n}), t \geq 0)$ obey an LDP in $D(R)$ for the J_1 topology and normalizing sequence b_n^2 with rate function

$$I(x) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty (\mu\dot{x}(t) - r)^2\, dt,$$

when x is absolutely continuous, $x(0) = 0$, and $I(x) = \infty$ otherwise. So the rate functions look similarly.

Lemma 2.1 and part (d) of Theorem 3.1 allow us to obtain LDPs for the other processes. For instance, we have the next result.

Corollary 3.1. *Under the conditions of Theorem 3.2, the following holds.*

(a) *The sequence* $\{\overline{W}_n, n \geq 1\}$ *obeys an LDP in* $D(R)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function*

$$I_W(w) = \frac{1}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(w(t) > 0)(\mu \dot{w}(t) - r)^2 \, dt + \frac{1(r > 0) r^2}{2(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(w(t) = 0) \, dt,$$

when w *is nonnegative and absolutely continuous with* $w(0) = 0$ *and* $I_W(w) = \infty$ *otherwise.*

(b) *The sequence* $\{\overline{H}_n, n \geq 1\}$ *obeys an LDP in* $D(R)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function*

$$I_H(h) = \frac{1}{2\mu(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(h(t) > 0)(\mu^2 \dot{h}(t) - r)^2 \, dt + \frac{1(r > 0) r^2}{2\mu(\sigma_A^2 + \sigma_S^2)} \int_0^\infty 1(h(t) = 0) \, dt,$$

when h *is nonnegative and absolutely continuous with* $h(0) = 0$ *and* $I_H(h) = \infty$ *otherwise.*

Proof. For the proof it suffices to observe that by part (d) of Theorem 3.1 and Lemma 2.1, $I_W(w) = I_Q(\mu w)$ and $I_H(h) = I_W(h \circ (\mu e))$.

We can also project even more to get LDPs for one-dimensional distributions. To illustrate, we give two examples. Denote $r' = -r\mu^{-2}$.

Corollary 3.2. *Let the conditions of Theorem 3.2 hold. Then*

(a) *the sequence* $\{\overline{H}_n(t), n \geq 1\}$ *for* $t \geq 0$ *obeys an LDP in* R_+ *for normalizing sequence* b_n^2 *with rate function*

$$I_{H(t)}(z) = \begin{cases} \dfrac{\mu^3}{2(\sigma_A^2 + \sigma_S^2)} \dfrac{(z + r't)^2}{t}, & \text{when } r' < 0 \text{ or } r' \geq 0, \ \dfrac{z}{r'} > t, \\ \dfrac{2\mu^3 r' z}{\sigma_A^2 + \sigma_S^2}, & \text{when } r' > 0, \ \dfrac{z}{r'} \leq t; \end{cases}$$

(b) *the sequence* $\{\overline{C}_n(t), n \geq 1\}$ *for* $t \geq 0$ *obeys an LDP in* R_- *for normalizing sequence* b_n^2 *with rate function*

$$I_{C(t)}(z) = \begin{cases} \dfrac{\mu^2}{2(\sigma_A^2 + \sigma_S^2)} \dfrac{(z - r\mu^{-1}t)^2}{t}, & \text{when } r > 0 \text{ or } r \leq 0, \ \mu\dfrac{-z}{r} > t, \\ \dfrac{-2\mu r z}{\sigma_A^2 + \sigma_S^2}, & \text{when } r < 0, \ \mu\dfrac{-z}{r} \leq t. \end{cases}$$

Remark 3.5. Note that in "the ergodic case" $r < 0$, the rate function for $\{\overline{C}_n(t), n \geq 1\}$ is the same as for the arrived work $\{(V_n \circ A_n(t) - nt)/(b_n\sqrt{n}), n \geq 1\}$ which follows by Remark 3.4.

Remark 3.6. We do not know an explicit expression for I_D and I_L.

We end the section by showing, analogously to diffusion approximation results, that the LDPs for the processes of waiting and departure times can be established directly without invoking LDPs

for continuous-time processes, and that for the ergodic $GI/GI/1$ queue an LDP holds for stationary waiting times as well (cf. Prohorov [12]). Let us denote by $u_{n,i}, i \geq 1$, the time between the ith and $(i+1)$th arrivals and by $v_{n,i}, i \geq 1$, the service time of the ith customer in the nth system. The associated partial-sum processes $U'_n = (U'_n(k), k = 0, 1, 2, \ldots)$ and $V_n = (V_n(k), k = 0, 1, 2, \ldots)$ are given by

$$U'_n(k) = \sum_{i=1}^{k} u_{n,i},\ U'_n(0) = 0,\quad V_n(k) = \sum_{i=1}^{k} v_{n,i},\ V_n(0) = 0, \tag{3.32}$$

so that, as above, $V_n(k)$ is the cumulative service time of the first k customers. The obvious equations for waiting and departure times are

$$H_n(k+1) = V_n(k) - U'_n(k) - \min_{0 \leq i \leq k} (V_n(i) - U'_n(i)), \tag{3.33}$$

$$L_n(k+1) = U'_n(k) + H_n(k+1) + v_{n,k+1}. \tag{3.34}$$

Let

$$\tilde{U}'_n = (\tilde{U}'_n(t), t \geq 0),\ \tilde{U}'_n(t) = \frac{1}{b_n\sqrt{n}}(U'_n(\lfloor nt \rfloor) - \lambda_n^{-1} nt). \tag{3.35}$$

Recall that if the nth queue is a $GI/GI/1$ queue with $\lambda_n < \mu_n$, then the waiting times $H_n(k)$ converge in distribution as $k \to \infty$ to a proper random variable (see, e.g., Borovkov [1]). We denote the latter by H_n^0 and let $\overline{H}_n^0 = H_n^0/(b_n\sqrt{n})$.

Theorem 3.3. *Let (3.15) hold.*

(a) *Assume that* $\{(\tilde{U}'_n, \overline{V}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for one of the topologies* J_1, M_1 *or* M'_1 *and normalizing sequence* b_n^2 *with rate function* $I_{U,V}(u,v)$.

Then $\{\overline{H}_n, n \geq 1\}$ *obeys an LDP in* $D(R)$ *for the same topology and normalizing sequence* b_n^2 *with rate function*

$$I_H(h) = \inf_{\substack{u,v \in D(R^2): \\ h = \mathcal{R}(v-u-r'e)}} I_{U,V}(u,v)\ .$$

If, in addition, $I_{U,V}(u,v)$ *is infinite when* v *is either discontinuous or not equal to* 0 *at* 0, *then* $\{(\overline{H}_n, \overline{L}_n), n \geq 1\}$ *obeys an LDP in* $D(R^2)$ *for the same topology and normalizing sequence* b_n^2 *with rate function*

$$I_{H,L}(h,l) = \inf_{\substack{u,v \in D(R^2): \\ h = \mathcal{R}(v-u-r'e), \\ l = u+h+r'e}} I_{A,V}(a,v)\ .$$

(b) *Consider a sequence of $GI/GI/1$ queues for which the conditions of Remark 3.3 hold. Assume that $r < 0$. Then the sequence $\{\overline{H}_n^0, n \geq 1\}$ obeys an LDP in R_+ for the normalizing sequence b_n^2 with rate function*

$$I_{H^0}(z) = \frac{2r'z}{\sigma_U^2 + \sigma_V^2}.$$

Proof. We begin with part (a). For the part related to H_n, we use that by (2.4), (3.33), (3.35), (3.8), and (3.13)

$$\overline{H}_n = \mathcal{R}\left(\overline{V}_n - \tilde{U}_n' - (\lambda_n^{-1} - \mu_n^{-1})\frac{\sqrt{n}}{b_n}e\right).$$

For the second claim, we use that by (3.34), (3.35), (3.14), (3.13), and (3.8)

$$\overline{L}_n(t) = \tilde{U}_n'(t) + (\lambda_n^{-1} - \mu_n^{-1})\frac{\sqrt{n}}{b_n}t + \overline{H}_n(t) + \frac{v_{n,\lfloor nt\rfloor+1}}{b_n\sqrt{n}},$$

the fact that by the hypotheses and Lemma 2.3 $\sup_{s\leq t} v_{n,\lfloor ns\rfloor+1}/(b_n\sqrt{n}) \overset{P^{1/a_n}}{\to} 0$, and Lemma 4.2(b) in [17].

We now prove part (b). The argument is borrowed from the corresponding proofs of diffusion approximation results [12]. Since H_n^0 is distributed as $\sup_{k\geq 0}(V_n(k) - U_n'(k))$ [1], we have, for a Borel subset A of R_+,

$$\left|P(\overline{H}_n^0 \in A) - P\left(\frac{1}{b_n\sqrt{n}}\sup_{0\leq k\leq\lfloor nt\rfloor}(V_n(k) - U_n'(k)) \in A\right)\right| \leq P(\sup_{k>\lfloor nt\rfloor}(V_n(k) - U_n'(k)) \geq 0).$$

Since $\sup_{0\leq k\leq\lfloor nt\rfloor}(V_n(k) - U_n'(k))$ coincides in distribution with $H_n(\lfloor nt\rfloor + 1)$, and $\{H_n(\lfloor nt\rfloor + 1)/(b_n\sqrt{n}), n \geq 1\}$ by Corollary 3.2(a) obeys an LDP with the rate function $I_{H(t)}(z)$ for which $\lim_{t\to\infty}\inf_{z\in A} I_{H(t)}(z) = \inf_{z\in A} I_{H^0}(z)$, when $A = [a,b], [a,\infty), (-\infty, b], (a,b)$, Lemma 2.2 implies that the required would follow by

$$\lim_{t\to\infty}\overline{\lim_{n\to\infty}} P^{1/a_n}(\sup_{k>\lfloor nt\rfloor}(V_n(k) - U_n'(k)) \geq 0) = 0, \tag{3.36}$$

where, as above, $a_n = b_n^2$.

Denoting $\delta_n = E(u_{n,1} - v_{n,1})$ and $\xi_{n,i} = v_{n,i} - u_{n,i} + \delta_n$, we have, since $\delta_n > 0$,

$$\begin{aligned}
P(\sup_{k>\lfloor nt\rfloor}(V_n(k) - U_n'(k)) \geq 0) &\leq \sum_{l=\lfloor\log_2(nt)\rfloor}^{\infty} P\left(\max_{2^l+1\leq k\leq 2^{l+1}}\left(\sum_{i=1}^{k}\xi_{n,i} - k\delta_n\right) \geq 0\right) \\
&\leq \sum_{l=\lfloor\log_2(nt)\rfloor}^{\infty} P\left(\sum_{i=1}^{2^l}\xi_{n,i} \geq 2^{l-1}\delta_n\right) + \sum_{l=\lfloor\log_2(nt)\rfloor}^{\infty} P\left(\max_{1\leq k\leq 2^l}\sum_{i=1}^{k}\xi_{n,i} \geq 2^{l-1}\delta_n\right) \\
&\leq 2\sum_{l=\lfloor\log_2(nt)\rfloor}^{\infty} P\left(\max_{1\leq k\leq 2^l}\sum_{i=1}^{k}\xi_{n,i} \geq 2^{l-1}\delta_n\right).
\end{aligned}$$

Limit (3.36) now follows by Lemma A.1 in the appendix and the near-heavy traffic condition $\frac{\sqrt{n}}{b_n}\delta_n \to r' > 0$ as $n \to \infty$. The theorem is proved.

4. Moderate Deviations for Queueing Networks in Near-Heavy Traffic

We now extend some of the above results to the queueing-networks set-up. Our results here are in the spirit of Reiman [19]. We consider a sequence of networks indexed by n. The nth network has a homogeneous customer population and consists of K FIFO single server stations. The network is open in that customers arrive from outside and eventually leave. Let $A_{n,k}(t), 1 \le k \le K$, be the cumulative number of customers who arrived at station k from outside the network during the interval $[0,t]$, and let $S_{n,k}(t), 1 \le k \le K$, be the cumulative number of customers who are served at station k for the first t units of busy time of that station. We call $A_n = (A_{n,k}, 1 \le k \le K)$, where $A_{n,k} = (A_{n,k}(t), t \ge 0)$, and $S_n = (S_{n,k}, 1 \le k \le K)$, where $S_{n,k} = (S_{n,k}(t), t \ge 0)$, the arrival process and service process respectively (note that some of the entries in A_n may equal zero). We associate with the stations of the network the processes $\Phi_{n,k} = (\Phi_{n,kl}, 1 \le l \le K)$, $1 \le k \le K$, where $\Phi_{n,kl} = (\Phi_{n,kl}(m), m = 1,2,\ldots)$, and $\Phi_{n,kl}(m)$ denotes the cumulative number of customers among the first m customers who depart station k that go directly to station l. The process $\Phi_n = (\Phi_{n,kl}, 1 \le k,l \le K)$ is referred to as the routing process. We consider the processes $A_{n,k}$, $S_{n,k}$ and $\Phi_{n,k}$ as random elements of the respective Skorohod spaces $D(R)$, $D(R)$ and $D(R^K)$; accordingly, A_n, S_n and Φ_n are regarded to be random elements of $D(R^K)$, $D(R^K)$ and $D(R^{K\times K})$, respectively.

We next introduce normalized and time-scaled versions of the arrival process, service process and routing process. Let $\lambda_{n,k} \ge 0$, $\mu_{n,k} \ge 0$, and $p_{kl} \in [0,1]$, $1 \le k \le K, 1 \le l \le K$. Define

$$\overline{A}_{n,k}(t) = \frac{A_{n,k}(nt) - \lambda_{n,k}nt}{b_n\sqrt{n}}\ ,\ \overline{S}_{n,k}(t) = \frac{S_{n,k}(nt) - \mu_{n,k}nt}{b_n\sqrt{n}}\ ,\ \overline{\Phi}_{n,kl}(t) = \frac{\Phi_{n,kl}(nt) - p_{kl}nt}{b_n\sqrt{n}}\ , \quad (4.1)$$

where as above $b_n \to \infty$ and $b_n/\sqrt{n} \to 0$, and let $\overline{A}_n = (\overline{A}_{n,k}, 1 \le k \le K)$, $\overline{S}_n = (\overline{S}_{n,k}, 1 \le k \le K)$, $\overline{\Phi}_{n,k} = (\overline{\Phi}_{n,kl}, 1 \le l \le K), 1 \le k \le K$, and $\overline{\Phi}_n = (\overline{\Phi}_{n,kl}, 1 \le k,l \le K)$. Again the latter processes are considered as random elements of $D(R^K)$, $D(R^K)$, $D(R^K)$, and $D(R^{K\times K})$, respectively. Also we denote $\lambda_n = (\lambda_{n,k}, 1 \le k \le K)$, $\mu_n = (\mu_{n,k}, 1 \le k \le K)$ and $P = (p_{kl}, 1 \le k \le K, 1 \le l \le K)$. The first two vectors as well as other elements of R^K are regarded to be column-vectors.

In analogy with the hypotheses of Section 3, we assume that $\lambda_n \to \lambda \equiv (\hat{\lambda}_1, \ldots, \hat{\lambda}_K)$ and $\mu_n \to \mu \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_K)$ as $n \to \infty$, where μ is componentwise positive, and that the near-heavy traffic condition holds: for some $r \in R^K$,

$$\frac{\sqrt{n}}{b_n}(\lambda_n - (I - P^T)\mu_n) \to r \quad \text{as } n \to \infty, \quad (4.2)$$

in particular,

$$\lambda = (I - P^T)\mu. \tag{4.3}$$

(As above, ^{T} denotes taking the transpose of either a matrix or a vector.) We also assume that the spectral radius of the matrix P is less than unity.

Our main concern here is the queue-length process $Q_n = (Q_{n,k}, 1 \le k \le K)$, where $Q_{n,k} = (Q_{n,k}(t), t \ge 0)$, with $Q_{n,k}(t)$ denoting the number of customers at station k at time t. Other related processes can be treated analogously to Section 3. The associated normalized and time-scaled process $\overline{Q}_n = (\overline{Q}_{n,k}, 1 \le k \le K)$ is defined by

$$\overline{Q}_{n,k}(t) = \frac{Q_{n,k}(nt)}{b_n\sqrt{n}}. \tag{4.4}$$

We fix some notation. If $x \in D(R^K)$ has componentwise nondecreasing nonnegative paths, then for $y \in D(R^K)$ we denote $y \circ x = ((y_k \circ x_k(t), 1 \le k \le K), t \ge 0)$, accordingly, if $\phi(t) = (\phi_{kl}(t), 1 \le k, l \le K) \in R^{K\times K}$, then $\phi \circ x(t) \equiv (\phi_{kl} \circ x_k(t), 1 \le k, l \le K)$. For a vector $\alpha = (\alpha_1, \ldots, \alpha_K) \in R^K$, we denote $\alpha e = ((\alpha_1 t, \ldots, \alpha_k t), t \ge 0)$. For a subset J of $\{1, 2, \ldots, K\}$, we set $F_J = \{\alpha = (\alpha_1, \ldots, \alpha_K) \in R_+^K : \alpha_k = 0, k \in J, \alpha_k > 0, k \notin J\}$ and $\overline{F}_J = \{\alpha = (\alpha_1, \ldots, \alpha_K) \in R_+^K : \alpha_k = 0, k \in J\}$; $\mathbf{1}_J$ is the K-vector with entries from J equal to 1 and the rest of the entries equal to 0; J^c denotes the complement of J. We also denote: R_+^0 is the interior of R_+, $\mathbf{1}$ is the K-vector with all the components equal to 1, $\mathcal{K}$ is the set of all the subsets of $\{1, 2, \ldots, K\}$ excluding the empty set. For vectors $\alpha = (\alpha_1, \ldots, \alpha_K) \in R^K$ and $\alpha' = (\alpha'_1, \ldots, \alpha'_K) \in R^K$, we denote $\alpha \otimes \alpha' = (\alpha_1\alpha'_1, \ldots, \alpha_K\alpha'_K) \in R^K$.

Theorem 4.1. *Let $Q_{n,k}(0) = 0, 1 \le k \le K, n \ge 1$, and the near-heavy traffic condition (4.2) hold.*

(a) *Assume that the sequence $\{(\overline{A}_n, \overline{S}_n, \overline{\Phi}_n), n \ge 1\}$ obeys an LDP in $D(R^K \times R^K \times R^{K\times K})$ for one of the J_1, M_1 or M_1' topologies and normalizing sequence b_n^2 with rate function $I_{A,S,\Phi}(a, s, \phi)$. Then $\{\overline{Q}_n, n \ge 1\}$ obeys an LDP in $D(R^K)$ for the same topology and normalizing sequence b_n^2 with rate function*

$$I_Q(q) = \inf_{\substack{a,s,\phi\in D(R^K\times R^K\times R^{K\times K}):\\ q=\mathcal{R}_P(a+(\phi\circ\mu e)^T\cdot\mathbf{1}-(I-P^T)s+re)}} I_{A,S,\Phi}(a, s, \phi)\ .$$

(b) *Assume, in addition, that $I_{A,S,\Phi}$ has the following form: for $a = (a_1, \ldots, a_K) \in D(R^K)$, $s = (s_1, \ldots, s_K) \in D(R^K)$ and $\phi = (\phi_1, \ldots, \phi_K) \in D(R^{K\times K})$,*

$$I_{A,S,\Phi}(a, s, \phi) = \sum_{k=1}^K I_{A_k}(a_k) + \sum_{k=1}^K I_{S_k}(s_k) + \sum_{k=1}^K I_{\Phi_k}(\phi_k),$$

where

$$I_{A_k}(a_k) = \frac{1}{2\sigma_{a,k}^2} \int_0^\infty \dot{a}_k(t)^2 dt$$

for a_k *absolutely continuous with* $a_k(0) = 0$ *and* $I_{A_k}(a_k) = \infty$ *otherwise,*

$$I_{S_k}(s_k) = \frac{1}{2\sigma_{s,k}^2} \int_0^\infty \dot{s}_k(t)^2 dt$$

for s_k *absolutely continuous with* $s_k(0) = 0$ *and* $I_{S_k}(s_k) = \infty$ *otherwise, and*

$$I_{\Phi_k}(\phi_k) = \int_0^\infty \sup_{\lambda \in R^k} \left(\lambda^T \dot{\phi}_k(t) - \frac{1}{2} \lambda^T \Sigma_{\Phi,k} \lambda \right) dt,$$

for $\phi_k = (\phi_{kl}, 1 \le l \le K)$ *absolutely continuous with* $\phi_{kl}(0) = 0$ *and* $I_{\Phi_k}(\phi_k) = \infty$ *otherwise, where* $\Sigma_{\Phi,k}, 1 \le k \le K$, *are symmetric nonnegative-definite* $K \times K$ *matrices.*

Assume that the symmetric nonnegative-definite $K \times K$ *matrix* Γ *defined by*

$$\Gamma = \mathrm{diag}(\sigma_{a,1}^2, \ldots, \sigma_{a,K}^2) + (I - P^T)\,\mathrm{diag}(\sigma_{s,1}^2, \ldots, \sigma_{s,K}^2)\,(I - P) + \sum_{k=1}^K \hat{\mu}_k \Sigma_{\Phi,k}$$

is positive definite.

Then $\{Q_n, n \ge 1\}$ *obeys an LDP in* $D(R^K)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function*

$$I_Q(q) = \frac{1}{2} \int_0^\infty 1(q(t) \in R_+^0)(\dot{q}(t) - r)^T \Gamma^{-1} (\dot{q}(t) - r)\, dt$$

$$+ \sum_{J \in \mathcal{K}} \frac{1}{2} \int_0^\infty 1(q(t) \in F_J) \inf_{y \in \overline{F}_{J^c}} (\dot{q}(t) \otimes \mathbf{1}_{J^c} - r - (I - P^T)y)^T \Gamma^{-1} (\dot{q}(t) \otimes \mathbf{1}_{J^c} - r - (I - P^T)y)\, dt,$$

when q is absolutely continuous with $q(0) = 0$ and $I_Q(q) = \infty$ otherwise.

(c) *Assume that the processes* $\overline{A}_{n,k}, 1 \le k \le K$, $\overline{S}_{n,k}, 1 \le k \le K$, *and* $\overline{\Phi}_{n,k}, 1 \le k \le K$, *are mutually independent for each* n. *Assume that the processes* $\overline{A}_{n,k}$ *and* $\overline{S}_{n,k}$ *are renewal processes and let* $\hat{u}_{n,k}$ *denote the generic exogenous interarrival time and* $\hat{v}_{n,k}$, *the generic service time in station* k. *Let the stations be indexed so that, for some* K', $\hat{\lambda}_i > 0$ *when* $1 \le i \le K'$, *and* $\hat{\lambda}_i = 0$ *when* $K' + 1 \le i \le K$. *Let*

$$E\hat{u}_{n,k} \to \hat{\lambda}_k^{-1}, \ \mathrm{Var}\ \hat{u}_{n,k} \to \sigma_{u,k}^2, \ 1 \le k \le K',$$
$$E\hat{v}_{n,k} \to \hat{\mu}_k^{-1}, \ \mathrm{Var}\ \hat{v}_{n,k} \to \sigma_{v,k}^2, \ 1 \le k \le K,$$

and either one of the following conditions be met:

(i) $\sup_n E(\hat{u}_{n,k})^{2+\epsilon} < \infty, 1 \le k \le K'$, $\sup_n E(\hat{v}_{n,k})^{2+\epsilon} < \infty, 1 \le k \le K$, *for some* $\epsilon > 0$ *and* $\sqrt{\log n}/b_n \to \infty$;

(ii) $\sup_n E\exp(\alpha(\hat{u}_{n,k})^\beta) < \infty$, $1 \le k \le K'$, $\sup_n E\exp(\alpha(\hat{v}_{n,k})^\beta) < \infty, 1 \le k \le K$, *for some* $\alpha > 0, 0 < \beta \le 1$ *and* $n^{\beta/2}/b_n^{2-\beta} \to \infty$.

If, in addition, the routing mechanism does not depend on n *and is i.i.d. at each station with* p_{kl} *being the probability of going directly from station* k *to station* l*, then the conditions of part (b) hold with*

$$\sigma_{a,k}^2 = \sigma_{u,k}^2 \hat{\lambda}_k^3,\ 1 \le k \le K',\ \sigma_{a,k}^2 = 0,\ K'+1 \le k \le K,\ \sigma_{s,k}^2 = \sigma_{v,k}^2 \hat{\mu}_k^3,\ 1 \le k \le K,$$

$$(\Sigma_{\Phi,k})_{l,m} = \begin{cases} p_{kl}(1-p_{kl}), & \text{if } m = l, \\ -p_{kl}p_{km}, & \text{if } m \ne l, \end{cases} \quad 1 \le k \le K, 1 \le l \le K, 1 \le m \le K,$$

and

$$I_{\Phi_k}(\phi_k) = \frac{1}{2}\sum_{l=1}^{K} \int_0^\infty \frac{\dot{\phi}_{kl}(t)^2}{p_{kl}}\, dt,\ 1 \le k \le K,$$

for $\phi_k = (\phi_{kl}, 1 \le l \le K)$ *absolutely continuous with* $\phi_{kl}(0) = 0$ *and* $\sum_{l=1}^K \phi_{kl}(t) = 0$*, and* $I_{\Phi_k}(\phi_k) = \infty$ *otherwise.*

Remark 4.1. If the matrix Γ is degenerate, then the LDP in part (b) holds with the same rate function I_Q provided in its definition expressions of the form $\frac{1}{2}x^T\Gamma^{-1}x$, $x \in R^k$, are understood as $\sup_{\lambda \in R^k}(\lambda^T x - \frac{1}{2}\lambda^T\Gamma\lambda)$.

Proof of Theorem 4.1. The proof is a straightforward extension of the proof of Theorem 3.1 (cf., a similar argument in the proofs of corresponding weak convergence results in [19, 3]). In analogy with (3.4), (3.5) and (3.3), we have that for $1 \le k \le K$

$$Q_{n,k}(t) = A_{n,k}(t) + \sum_{l=1}^{K} \Phi_{n,lk} \circ D_{n,l}(t) - D_{n,k}(t),$$

where

$$D_{n,k}(t) = S_{n,k}\left(\int_0^t 1(Q_{n,k}(s) > 0)\, ds\right).$$

Introducing

$$\overline{C}'_{n,k}(t) = \int_0^t 1(\overline{Q}_{n,k}(s) > 0)\, ds,\ \overline{D}'_{n,k}(t) = \frac{D_{n,k}(nt)}{n},$$

we then have by (4.1) and (4.4) that

$$\begin{aligned} \overline{Q}_{n,k}(t) &= \overline{A}_{n,k}(t) + \sum_{l=1}^{K} \overline{\Phi}_{n,lk} \circ \overline{D}'_{n,l}(t) \\ &+ \sum_{l=1}^{K} p_{lk}\overline{S}_{n,l} \circ \overline{C}'_{n,l}(t) - \overline{S}_{n,k} \circ \overline{C}'_{n,k}(t) + \frac{\sqrt{n}}{b_n}(\lambda_{n,k} + \sum_{l=1}^{K} p_{lk}\mu_{n,l} - \mu_{n,k})t \\ &+ \frac{\sqrt{n}}{b_n}\left(\mu_{n,k}\int_0^t 1(\overline{Q}_{n,k}(s) = 0)\, ds - \sum_{l=1}^{K} p_{lk}\mu_{n,l}\int_0^t 1(\overline{Q}_{n,l}(s) = 0)\, ds\right) \end{aligned} \tag{4.5}$$

which implies that

$$\overline{Q}_n = \mathcal{R}_P(\overline{A}_n + (\overline{\Phi}_n \circ \overline{D}'_n)^T \cdot 1 - (I - P^T)\,\overline{S}_n \circ \overline{C}'_n + \frac{\sqrt{n}}{b_n}(\lambda_n + (P^T - I)\mu_n)e) \tag{4.6}$$

and hence

$$\frac{\sqrt{n}}{b_n}(I - P^T)\mu_n \otimes \overline{C}'_n = \overline{A}_n + (\overline{\Phi}_n \circ \overline{D}'_n)^T \cdot 1 - (I - P^T)\,\overline{S}_n \circ \overline{C}'_n + \frac{\sqrt{n}}{b_n}(\lambda_n + (P^T - I)\mu_n)e - \overline{Q}_n,$$

where $\overline{C}'_n(t) = (\overline{C}'_{n,k}(t), 1 \le k \le K)$ and $\overline{D}'_n(t) = (\overline{D}'_{n,k}(t), 1 \le k \le K)$. The Lipshitz property of $\mathcal{R}_P$, the LDP for $\{(\overline{A}_n, \overline{S}_n, \overline{\Phi}_n), n \ge 1\}$, (4.2), and the fact that $I - P^T$ is nonsingular yield by the argument of the proof of (3.23), since μ is componentwise positive,

$$\int_0^t 1(\overline{Q}_{n,k}(s) = 0)\, ds \overset{P^{1/a_n}}{\to} 0 \text{ as } n \to \infty,\ 1 \le k \le K,\ t > 0,$$

where again $a_n = b_n^2$, implying that

$$\overline{C}'_{n,k} \overset{P^{1/a_n}}{\to} e \text{ as } n \to \infty. \tag{4.7}$$

Then by Lemma 4.2(b) in [17]

$$\overline{D}'_n \overset{P^{1/n}}{\to} \mu e \tag{4.8}$$

after which Lemma 4.3 in [17] enables us to conclude that the sequence $\{(\overline{A}_n, \overline{S}_n \circ \overline{C}'_n, \overline{\Phi}_n \circ \overline{D}'_n), n \ge 1\}$ obeys an LDP in $D(R^K \times R^K \times R^{K\times K})$ with rate function $\overline{I}_{A,S,\Phi}$ given by the equality $\overline{I}_{A,S,\Phi}(a, s, \phi \circ \mu e) = I_{A,S,\Phi}(a, s, \phi)$. The claim of part (a) follows by (4.6) and the contraction principle.

Part (b) is a consequence of part (a) and Lemma 2.4. In more detail, we have by part (a), Lemma 3.3 in [15] and Lemma 2.4

$$\begin{aligned} I_Q(q) = &\inf_{\substack{(a,s,\phi)\in D(R^K\times R^K\times R^{K\times K}):\\ q=\mathcal{R}_P(a+(\phi\circ\mu e)^T 1-(I-P^T)s+re)}} \left(\sum_{k=1}^K I_{A_k}(a_k) + \sum_{k=1}^K I_{S_k}(s_k) + \sum_{k=1}^K I_{\Phi_k}(\phi_k)\right) \\ = &\int_0^\infty \inf_{\substack{(\alpha,\beta,\psi,\gamma)\in R^K\times R^K\times R^{K\times K}\times R_+^K:\\ \dot q(t)=\alpha+\psi^T\cdot\mu-(I-P^T)\beta+r+(I-P^T)\gamma,\\ \gamma_k q_k(t)=0, 1\le k\le K}} \left(\sum_{k=1}^K \frac{1}{2\sigma_{a,k}^2}\alpha_k^2 + \sum_{k=1}^K \frac{1}{2\sigma_{s,k}^2}\beta_k^2 \right. \\ &\left. + \sum_{k=1}^K \hat\mu_k \sup_{\lambda\in R^k}\left(\lambda^T\psi_k - \frac{1}{2}\lambda^T\,\Sigma_{\Phi,k}\,\lambda\right)\right) dt. \end{aligned}$$

By mean squares, the infimum in the integral over α, β and ψ, for γ fixed, equals $(\dot q(t) - r - (I - P^T)\gamma)^T\Gamma^{-1}(\dot q(t) - r - (I - P^T)\gamma)$. This completes the proof of (b).

The conditions of (c) imply the conditions of (b) by Lemma 6.1 in [17]. The theorem is proved.

Remark 4.2. Note that the matrix Γ in part (c) coincides with the covariance matrix in Reiman's result [19].

Remark 4.3. The rate function in part (b) is not as explicit as in Theorem 3.2 in that on the faces F_J we need to solve quadratic programming problems. It appears that generally this needs to be done numerically.

Remark 4.4. The contraction principle allows us to deduce that under the conditions of the theorem one-dimensional projections also obey LDPs. An open question is deriving explicitly the rate functions as in Corollary 3.2. It is not difficult to see that the optimal paths q must be piecewise linear. However, we can solve explicitly only the case $K = 2$ (cf., Ignatyuk, Malyshev and Scherbakov [6]).

We now apply Theorem 4.1 to obtain LDPs for waiting and sojourn times (cf., Reiman [19]). Let $W_{n,k}(t), 1 \le k \le K$, denote the virtual waiting time at station k at time t. Define $\overline{W}_{n,k}(t) = W_{n,k}(nt)/(b_n\sqrt{n})$ and let $\overline{W}_n = ((\overline{W}_{n,k}(t), 1 \le k \le K), t \ge 0)$. Next, for a vector $\mathbf{k} = (k_1, \ldots, k_l)$, where $k_i \in \{1, 2, \ldots, K\}$, let $A_{n,\mathbf{k}}(t)$ denote the number of customers with the routing $(k_1, k_2, \ldots, k_l)$ who have exogenously arrived by t and $Y_{n,\mathbf{k}}(m)$ denote the sojourn time of the mth exogenous customer with the routing $(k_1, k_2, \ldots, k_l)$, and let $\overline{Y}_{n,\mathbf{k}}(t) = Y_{n,\mathbf{k}}(\lfloor nt \rfloor + 1)/(b_n\sqrt{n})$, $\overline{Y}_{n,\mathbf{k}} = (\overline{Y}_{n,\mathbf{k}}(t), t \ge 0)$, $\overline{A}'_{n,\mathbf{k}} = (A_{n,\mathbf{k}}(nt)/n, t \ge 0)$.

Corollary 4.1. (a). *Assume that the sequence* $\{(\overline{A}_n, \overline{S}_n, \overline{\Phi}_n), n \ge 1\}$ *obeys an LDP in* $D(R^K \times R^K \times R^{K\times K})$ *for one of the* J_1 *or* M'_1 *topologies and normalizing sequence* b_n^2 *with rate function* $I_{A,S,\Phi}(a, s, \phi)$, *which, in the case of the* J_1 *topology, equals infinity unless* s *is continuous and equal to* 0 *at* 0. *Then the sequence* $\{(\overline{Q}_n, \overline{W}_n), n \ge 1\}$ *obeys an LDP in* $D(R^K \times R^K)$ *for the same topology and normalizing sequence* b_n^2 *with rate function* $I_{Q,W}(q, w)$ *such that* $q = \mu \otimes w$, *when* $I_{Q,W}(q, w) < \infty$. *In particular, the sequence* $\{\overline{W}_n, n \ge 1\}$ *obeys an LDP in* $D(R^K)$ *with rate function* $I_W(w) = I_Q(\mu \otimes w)$.

(b). *Assume, in addition, that the rate function* $I_{A,S,\Phi}(a, s, \phi)$ *equals infinity unless* a, s *and* ϕ *are both continuous and equal to* 0 *at* 0, *and*

$$\overline{A}'_{n,\mathbf{k}} \overset{P}{\to}{}^{1/a_n} \lambda_{\mathbf{k}} e \tag{4.9}$$

as $n \to \infty$, *for some* $\lambda_{\mathbf{k}} > 0$.

Then the sequence $\{(\overline{W}_n, \overline{Y}_{n,\mathbf{k}}), n \ge 1\}$ *obeys an LDP in* $D(R^K \times R)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function* $I_{W,Y}(w, y) = I_W(w)$, *when* $y \circ (\lambda_{\mathbf{k}} e) = \sum_{i=1}^{l} w_{k_i}$, *and* $I_{W,Y}(w, y) = \infty$ *otherwise, where* $w = (w_1, \ldots, w_K)$. *In particular, the sequence* $\{\overline{Y}_{n,\mathbf{k}}, n \ge$

1} *obeys an LDP in* $D(R)$ *for the* J_1 *topology and normalizing sequence* b_n^2 *with rate function* $I_{Y_{\mathbf{k}}}(y) = \inf_{w:y=\sum_{i=1}^{l} w_{k_i} \circ (\lambda_{\mathbf{k}}^{-1} e)} I_W(w)$.

Proof. The proof is much similar to that of Theorem 4.1. We only give a sketch. Let $V_{n,k}(m)$, for $k = 1, 2, \ldots, K$, and $m = 1, 2, \ldots$, denote the cumulative service time of the first m customers served at station k: $V_{n,k}(m) = \inf\{t \geq 0 : S_{n,k}(t) \geq m\}$, $V_{n,k}(0) = 0$, and let $E_{n,k}(t), t \geq 0$, denote the total number of arrivals to station k by time t: $E_{n,k}(t) = A_{n,k}(t) + \sum_{l=1}^{K} \Phi_{n,lk} \circ D_{n,l}(t)$. Introducing $\overline{V}_{n,k}(t) = V_{n,k}(\lfloor nt \rfloor)/(b_n\sqrt{n})$, $\overline{V}_n(t) = (\overline{V}_{n,1}(t), \ldots, \overline{V}_{n,K}(t))$, $\overline{V}_n = (\overline{V}_n(t), t \geq 0)$, $E'_{n,k}(t) = E_{n,k}(nt)/n$, $E'_n(t) = (E'_{n,1}(t), \ldots, E'_{n,K}(t))$ and $E'_n = (E'_n(t), t \geq 0)$, we have, in analogy with (4.5),

$$\mu_n \otimes \overline{W}_n(t) = \mu_n \otimes \overline{V}_n \circ E'_n(t) + \overline{A}_n(t) + \left(\overline{\Phi}_n^T \circ \overline{D}'_n(t)\right) \cdot \mathbf{1}$$
$$+P^T\left(\overline{S}_n \circ \overline{C}'_n(t)\right) + \frac{\sqrt{n}}{b_n}[\lambda_n + P^T\mu_n - \mu_n]t + \frac{\sqrt{n}}{b_n}[I - P^T](\mu_n \otimes \overline{C}'_n(t)).$$

In analogy with (4.8) and in view of (4.3), Lemma 4.2(b) in [17] implies that

$$E'_n \overset{P^{1/a_n}}{\to} \mu e. \tag{4.10}$$

Therefore, recalling (4.6), (4.8) and (4.7), we have that the sequence $\{(\overline{A}_n, \overline{S}_n, \overline{V}_n, \overline{\Phi}_n, \overline{Q}_n, \overline{W}_n), n \geq 1\}$ obeys an LDP in $D(R^K \times R^K \times R^K \times R^{K\times K} \times R^K \times R^K)$ with rate function $I_{A,S,V,\Phi,Q,W}(a, s, v, \phi, q, w)$ such that whenever it is finite the following equations hold

$$\mu \otimes w = \mathcal{R}_P(\mu \otimes v \circ (\mu e) + a + (\phi^T \circ (\mu e)) \cdot \mathbf{1} + P^T s + r),$$
$$s = -\mu \otimes v \circ (\mu e), \quad q = \mathcal{R}_P(a + (\phi^T \circ (\mu e)) \cdot \mathbf{1} - (I - P^T)s + r).$$

An application of Lemma 2.1 ends the proof of (a).

We now turn to (b). Note first that the argument of the proof of part (a) implies that the sequences $\{\overline{W}_n, n \geq 1\}$ and $\{\overline{V}_n, n \geq 1\}$ obey LDPs for the J_1 topology with rate functions which equal infinity both at discontinuous functions and functions not equal to 0 at 0.

Next, let us denote by $U_{n,\mathbf{k}}(m)$ the arrival time of the mth exogenous customer with the routing vector $\mathbf{k}$, by $T_{n,\mathbf{k},i}(m), 1 \leq i \leq l$, the time it arrives at the ith queue of its itinerary, by $H_{n,\mathbf{k},i}(m)$, the time it awaits service in the ith station and by $v_{n,\mathbf{k},i}(m)$, the time it is served in the ith station. We obviously have

$$T_{n,\mathbf{k},1}(m) = U_{n,\mathbf{k}}(m), \quad T_{n,\mathbf{k},i}(m) = T_{n,\mathbf{k},i-1}(m) + H_{n,\mathbf{k},i-1}(m) + v_{n,\mathbf{k},i-1}(m), \tag{4.11}$$

and

$$W_{n,k_i}(T_{n,\mathbf{k},i}(m)-) \leq H_{n,\mathbf{k},i}(m) \leq W_{n,k_i}(T_{n,\mathbf{k},i}(m)). \tag{4.12}$$

Inequalities (4.12) account for the fact that we make no assumptions about the mechanism of resolving conflicts between simultaneous arrivals. Next, it is easily seen that

$$\frac{1}{b_n\sqrt{n}}v_{n,\mathbf{k},i}(m) \leq \sup_{0\leq s\leq E'_{n,k_i}(T_{n,\mathbf{k},i}(m)/n)} |\Delta\overline{V}_{n,k_i}(s)|. \tag{4.13}$$

Since $E'_{n,k}(t) \overset{P^{1/a_n}}{\to} \hat{\mu}_k t$ by (4.10), $U_{n,\mathbf{k}}(\lfloor nt\rfloor+1)/n \overset{P^{1/a_n}}{\to} \lambda_{\mathbf{k}}^{-1}t$ by the assumption $\overline{A}'_{n,\mathbf{k}} \overset{P^{1/a_n}}{\to} \lambda_{\mathbf{k}}e$ and Lemma 4.2(c) in [17], $W_{n,k}(\lfloor nt\rfloor)/n \overset{P^{1/a_n}}{\to} 0$ by the LDP for $\{\overline{W}_n, n\geq 1\}$ and Lemma 4.2(c) in [17], and $\sup_{s\leq t}|\Delta\overline{V}_{n,k}(s)| \overset{P^{1/a_n}}{\to} 0$ by the fact that $\{\overline{V}_n, n\geq 1\}$ obeys an LDP with rate function that equals infinity both at discontinuous functions and functions not equal to 0 at 0 and Lemma 2.3, it follows from (4.11), (4.12) and (4.13) that, for $i=1,2,\ldots,l$,

$$\frac{1}{n}T_{n,\mathbf{k},i}(\lfloor nt\rfloor+1) \overset{P^{1/a_n}}{\to} \lambda_{\mathbf{k}}^{-1}t \tag{4.14}$$

and

$$\sup_{0\leq s\leq t}\frac{1}{b_n\sqrt{n}}v_{n,\mathbf{k},i}(\lfloor ns\rfloor+1) \overset{P^{1/a_n}}{\to} 0. \tag{4.15}$$

Let $\overline{H}_{n,\mathbf{k},i} = (H_{n,\mathbf{k},i}(\lfloor nt\rfloor+1)/(b_n\sqrt{n}), t\geq 0), 1\leq i\leq l$. The LDP for $\{\overline{W}_n, n\geq 1\}$, (4.14) and (4.12) imply, by Lemmas 4.1(c) and 4.3 in [17] and Lemma 2.3, that $\{(\overline{H}_{n,\mathbf{k},1},\ldots,\overline{H}_{n,\mathbf{k},l},\overline{W}_n), n\geq 1\}$ obeys an LDP in $D(R^l\times R^K)$ with rate function $I_{H_{\mathbf{k},1},\ldots,H_{\mathbf{k},l},W}(h_{\mathbf{k},1},\ldots,h_{\mathbf{k},l},w) = I_W(w)$, when $h_{\mathbf{k},i} = w_{k_i}\cdot(\lambda_{\mathbf{k}}^{-1}e)$, and equal to infinity otherwise. The proof is completed by noting that

$$\overline{Y}_{n,\mathbf{k}}(t) = \sum_{i=1}^{l}\overline{H}_{n,\mathbf{k},i}(t) + \sum_{i=1}^{l}\frac{1}{b_n\sqrt{n}}v_{n,\mathbf{k},i}(\lfloor nt\rfloor+1)$$

and using (4.15), Lemma 4.1(c) in [17] and the contraction principle. The corollary is proved.

Remark 4.5. If the routing mechansim is as described in part (c) of Theorem 4.1, then convergence (4.9) in part (b) holds with $\lambda_{\mathbf{k}} = p_{k_1k_2}\cdots p_{k_{l-1}k_l}\hat{\lambda}_{k_1}$. This follows by Theorem 6.3 in [18] and Lemma 4.2(b) in [17].

Acknowledgement. I am grateful to Ward Whitt for fruitful discussions and suggesting Theorem 3.3(a) and to Marty Reiman for valuable comments on the contents of the paper.

A. Appendix

We state and prove the lemma used in the proof of Theorem 3.3(b).

Lemma A.1. *Let $\{\xi_{n,i}, i\geq 1\}, n\geq 1$, be a triangular array of row-wise i.i.d. r.v. with zero mean. Let $b_n\to\infty$ as $n\to\infty$, and $\alpha>0$.*

(i) *If* $b_n/\sqrt{n} \to 0$ *as* $n \to \infty$ *and, for some* $\varepsilon > 0$, *we have* $\sup_n E|\xi_{n,1}|^{2+\varepsilon} < \infty$, *then there exist* n_0, $t_0 > 0$, $C_1 > 0$ *and* $C_2 > 0$ *such that, for all* $t \geq t_0$ *and* $n \geq n_0$,

$$P\left(\max_{1\leq k\leq \lfloor nt\rfloor} \frac{1}{b_n\sqrt{n}} \sum_{i=1}^{k} \xi_{n,i} > \alpha t\right) \leq \exp(-C_1 b_n^2\sqrt{t}) + C_2 \frac{b_n^{2+\varepsilon}}{n^{\varepsilon/2}} \frac{1}{t^{\varepsilon/2}}. \tag{A.1}$$

(ii) If for some $\gamma > 0$ *and* $\beta \in (0,1]$, *we have* $\sup_n E\exp(\gamma|\xi_{n,1}|^\beta) < \infty$ *and* $n^{\beta/2}/b_n^{2-\beta} \to \infty$ *as* $n \to \infty$, *then there exist* n_0', $t_0' > 0$, $C_1' > 0$ *and* $C_2' > 0$ *such that, for all* $t \geq t_0'$ *and* $n \geq n_0'$,

$$P\left(\max_{1\leq k\leq \lfloor nt\rfloor} \frac{1}{b_n\sqrt{n}} \sum_{i=1}^{k} \xi_{n,i} > \alpha t\right) \leq \exp(-C_1' b_n^2\sqrt{t}) + \exp(-C_2'(b_n\sqrt{nt})^\beta). \tag{A.2}$$

Proof. The argument uses the ideas of the proof of Example 7.2 in [14]. Let the conditions of (i) hold. We first prove that there exist $C_1 > 0$ and t_0 such that for $t \geq t_0$

$$P\left(\max_{1\leq k\leq \lfloor nt\rfloor} \frac{1}{b_n\sqrt{n}} \sum_{i=1}^{k} \xi_{n,i} 1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right) > \alpha t\right) \leq \exp(-C_1 b_n^2\sqrt{t}). \tag{A.3}$$

Denote $B = \sup_n E|\xi_{n,1}|^{2+\varepsilon}$ and let

$$\hat{\xi}_{n,i} = \frac{b_n}{\sqrt{n}}\left(\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right) - E\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right)\right).$$

By Doob's inequality (see, e.g., Liptser and Shiryaev [9, Theorem 1.9.1]), for $\lambda > 0$,

$$P\left(\max_{1\leq k\leq \lfloor nt\rfloor} \frac{1}{b_n\sqrt{n}} \sum_{i=1}^{k} \left(\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right) - E\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right)\right) > \frac{\alpha t}{2}\right)$$
$$\leq \frac{\left(Ee^{2\lambda\hat{\xi}_{n,1}}\right)^{\lfloor nt\rfloor}}{e^{\lambda b_n^2 \alpha t}}. \tag{A.4}$$

Since $E\hat{\xi}_{n,1} = 0$, $|\hat{\xi}_{n,1}| \leq 2\sqrt{t}$ and $E\hat{\xi}_{n,1}^2 \leq E\xi_{n,1}^2 b_n^2/n$, it follows that

$$Ee^{2\lambda\hat{\xi}_{n,1}} \leq 1 + 2\lambda^2 e^{4\lambda\sqrt{t}} E\hat{\xi}_{n,1}^2 \leq 1 + 2\lambda^2 e^{4\lambda\sqrt{t}} \frac{b_n^2}{n} B,$$

so $\left(Ee^{2\lambda\hat{\xi}_{n,1}}\right)^{\lfloor nt\rfloor} \leq \exp(2\lambda^2 e^{4\lambda\sqrt{t}} B t b_n^2)$. Choosing in (A.4) $\lambda = 1/\sqrt{t}$, $t_0 = (4e^4 B/\alpha)^2$ and $C_1 = \alpha/2$, we get

$$P\left(\max_{1\leq k\leq \lfloor nt\rfloor} \frac{1}{b_n\sqrt{n}} \sum_{i=1}^{k} \left(\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right) - E\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}| \leq \sqrt{t}\right)\right) > \frac{\alpha t}{2}\right) \leq \exp(-C_1 b_n^2\sqrt{t}). \tag{A.5}$$

Now note that, since $E\xi_{n,1} = 0$, by the Chebyshev inequality,

$$\left|E\xi_{n,1}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,1}| \leq \sqrt{t}\right)\right| = \left|E\xi_{n,1}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,1}| > \sqrt{t}\right)\right| \leq \frac{b_n^{1+\varepsilon}}{n^{(1+\varepsilon)/2}} \frac{B}{t^{(1+\varepsilon)/2}},$$

hence

$$\frac{\lfloor nt\rfloor}{b_n\sqrt{n}}\left|E\xi_{n,1}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,1}|\le a\right)\right|\le\frac{b_n^{\varepsilon}}{n^{\varepsilon/2}}Bt^{(1-\varepsilon)/2},$$

so, by the fact that $\sqrt{n}/b_n\to\infty$ as $n\to\infty$, for all n large enough and $t\ge t_0$,

$$P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|\le\sqrt{t}\right)>\alpha t\right)$$
$$\le P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\left(\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|\le a\right)-E\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|\le a\right)\right)>\frac{\alpha t}{2}\right),$$

which together with (A.5) proves (A.3).

Estimate (A.1) now follows by (A.3), the inequalities

$$P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\xi_{n,i}>\alpha t\right)\le P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|\le\sqrt{t}\right)>\alpha t\right)$$
$$+P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)$$

and

$$P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)\le\lfloor nt\rfloor P\left(\frac{b_n}{\sqrt{n}}|\xi_{n,1}|>\sqrt{t}\right)\le\lfloor nt\rfloor\frac{b_n^{2+\varepsilon}}{n^{1+\varepsilon/2}}\frac{B}{t^{1+\varepsilon/2}}.$$

Part (i) is proved.

For part (ii), we write

$$P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\xi_{n,i}>\alpha t\right)\le P\left(\max_{1\le k\le\lfloor nt\rfloor}\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{k}\xi_{n,i}1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|\le\sqrt{t}\right)>\frac{\alpha t}{2}\right)$$
$$+P\left(\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}|\xi_{n,i}|1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)>\frac{\alpha t}{2}\right).\qquad(A.7)$$

Noting that the conditions of part (ii) imply the conditions of part (i), we estimate the first term on the right with the help of (A.3). For the second, we use the inequality

$$P\left(\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}|\xi_{n,i}|1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)>\frac{\alpha t}{2}\right)\le P\left(\frac{1}{b_n\sqrt{n}}\max_{1\le i\le\lfloor nt\rfloor}|\xi_{n,i}|>\sqrt{t}\right)$$
$$+P\left(\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}|\xi_{n,i}|1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)1\left(\frac{1}{b_n\sqrt{n}}|\xi_{n,i}|\le\sqrt{t}\right)>\frac{\alpha t}{2}\right).\qquad(A.8)$$

We first work with the second probability on the right. We have, for $\lambda>0$, by the Chebyshev inequality,

$$P\left(\frac{1}{b_n\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}|\xi_{n,i}|1\left(\frac{b_n}{\sqrt{n}}|\xi_{n,i}|>\sqrt{t}\right)1\left(\frac{1}{b_n\sqrt{n}}|\xi_{n,i}|\le\sqrt{t}\right)>\frac{\alpha t}{2}\right)$$

$$\leq \left(E \exp\left(2\lambda \frac{b_n}{\sqrt{n}} |\xi_{n,1}| 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}| > \sqrt{t} \right) 1\left(\frac{1}{b_n\sqrt{n}} |\xi_{n,1}| \leq \sqrt{t} \right) \right) \right)^{\lfloor nt \rfloor} \exp(-\lambda b_n^2 \alpha t)$$

$$\leq \exp\left(ntE \exp\left(2\lambda \frac{b_n}{\sqrt{n}} |\xi_{n,1}| \right) 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}| > \sqrt{t} \right) 1\left(\frac{1}{b_n\sqrt{n}} |\xi_{n,1}| \leq \sqrt{t} \right) - \lambda b_n^2 \alpha t \right). \qquad \text{(A.9)}$$

Next, for $0 < \beta < 1, c > 0, \lambda c \leq \gamma/2$,

$$\begin{aligned}
&E \exp\left(2\lambda \frac{b_n}{\sqrt{n}} |\xi_{n,1}| \right) 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}| > \sqrt{t} \right) 1\left(\frac{1}{b_n\sqrt{n}} |\xi_{n,1}| \leq \sqrt{t} \right) \\
&\leq E \exp\left(2\lambda \frac{b_n}{\sqrt{n}} |\xi_{n,1}| \right) 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}|^{1-\beta} > c \right) 1\left(\frac{1}{b_n\sqrt{n}} |\xi_{n,1}| \leq \sqrt{t} \right) \\
&\quad + E \exp\left(2\lambda \frac{b_n}{\sqrt{n}} |\xi_{n,1}| \right) 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}|^{1-\beta} \leq c \right) 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,1}| > \sqrt{t} \right) \\
&\leq \exp\left(2\lambda b_n^2 \sqrt{t} - \gamma \left(\frac{c\sqrt{n}}{b_n} \right)^{\frac{\beta}{1-\beta}} E \exp\left(\gamma |\xi_{n,1}|^\beta \right) \right) \\
&\quad + E \exp\left(\left(2\lambda c + \frac{\gamma}{2} \right) |\xi_{n,1}|^\beta \right) \exp\left(-\frac{\gamma}{2} \left(\frac{\sqrt{nt}}{b_n} \right)^\beta \right). \qquad \text{(A.10)}
\end{aligned}$$

Taking $\lambda = 1/(2\sqrt{t})$ and $c = \gamma\sqrt{t}/2$, it is not difficult to see by (A.9), (A.10) and the condition $n^{\beta/2}/b_n^{2-\beta} \to 0$ as $n \to \infty$, that, for all n and t large enough,

$$P\left(\frac{1}{b_n\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} |\xi_{n,i}| 1\left(\frac{b_n}{\sqrt{n}} |\xi_{n,i}| > \sqrt{t} \right) 1\left(\frac{1}{b_n\sqrt{n}} |\xi_{n,i}| \leq \sqrt{t} \right) > \frac{\alpha t}{2} \right) \leq \exp\left(-C_1'' b_n^2 \sqrt{t} \right).$$

By a similar argument, this bound is seen to hold for $\beta = 1$ as well.

Finally, the first term on the right of (A.8) is estimated as

$$P\left(\frac{1}{b_n\sqrt{n}} \max_{1 \leq i \lfloor nt \rfloor} |\xi_{n,i}| > \sqrt{t} \right) \leq nt \frac{E e^{\gamma |\xi_{n,1}|^\beta}}{e^{\gamma (b_n\sqrt{nt})^\beta}} \leq \exp(-C_2'(b_n\sqrt{nt})^\beta).$$

Substituting the estimates into (A.7) finishes the proof of (ii). The lemma is proved.

References

[1] A.A. Borovkov, *Stochastic processes in queueing theory* (in Russian: Nauka, 1972, English translation: Springer, 1976).

[2] H. Chen and A. Mandelbaum, Leontief systems, RBV's and RBM's, in: *Proc. Imperial College Workshop on Applied Stochastic Processes*, eds. M.H.A. Davis and R.J. Elliot (Gordon and Breach, London, 1991).

[3] H. Chen and W. Whitt, Diffusion Approximations for Open Queueing Networks with Service Interruptions, Queueing Systems 13(1993) 335–359.

[4] J.M. Harrison and M.I. Reiman, Reflected Brownian Motion on an Orthant, Ann. Prob. 9(1981) 302–308.

[5] D.L. Iglehart and W. Whitt, Multiple Channel Queues in Heavy Traffic, I and II, Adv. Appl. Prob. 2(1970) 150–177 and 355–369.

[6] I.A. Ignatyuk, V. Malyshev and V.V. Scherbakov, Boundary effects in large deviation problems, Russ. Math. Surv., 1994, v. 49, no. 2, pp. 41–99.

[7] J.F.C. Kingman, On queues in heavy traffic, J. Roy. Statist. Soc. B24(1962) 383–392.

[8] T. Lindvall, Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0,\infty)$, J. Appl. Prob. 10(1973) 109–121.

[9] R.Sh. Liptser and A.N. Shiryaev, *Theory of Martingales* (Kluwer, Dordrecht, 1989).

[10] A. Mandelbaum, The Dynamic Complementarity Problem, unpublished manuscript (1989).

[11] J.L. Pomarede, A Unified Approach Via Graphs to Skorohod's Topologies on the Function Space D, Ph.D. dissertation, Department of Statistics, Yale University (1976).

[12] Yu.V. Prohorov, Transient phenomena in queueing processes (in Russian), Lit. Mat. Rink. 3(1963) 199–206.

[13] A. Puhalskii, On Functional Principle of Large Deviations, in: *New Trends in Probability and Statistics*, v. 1, eds. V. Sazonov and T. Shervashidze (VSP/Mokslas, 1991), pp. 198–218.

[14] A. Puhalskii, Large Deviations of Semimartingales Via Convergence of the Predictable Characteristics, Stochastics 49(1994) 27–85.

[15] A. Puhalskii, Large Deviation Analysis of the Single Server Queue, Queueing Systems 21(1995) 5–66.

[16] A. Puhalskii, Large Deviations of Semimartingales: a Maxingale Problem Approach. I. Limits as Solutions to a Maxingale Problem, preprint (1995).

[17] A. Puhalskii and W. Whitt, Functional Large Deviation Principles for First Passage Time Processes. Ann. Appl. Prob., to appear.

[18] A. Puhalskii and W. Whitt, Functional Large Deviation Principles for Waiting and Departure Processes (submitted).

[19] M.I. Reiman, Open Queueing Networks in Heavy Traffic, Math. Oper. Res. 9(1984) 441–458.

[20] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis* (Chapman and Hall, London, 1995).

[21] A.V. Skorohod, Limit Theorems for Stochastic Processes, Th. Prob. Appl. 1(1956) 261–290.

[22] S.R.S. Varadhan, *Large Deviations and Applications* (SIAM, Philadelphia, 1984).

[23] A.D. Wentzell, *Limit Theorems on Large Deviations for Markov Random Processes* (Nauka, Moscow, 1986, in Russian, English translation: Reidel, Dordrecht, 1989)

[24] W. Whitt, Some Useful Functions for Functional Limit Theorems, Math. Oper. Res. 1(1980) 67–85.